A Review of

Paul T. Decker Daniel P. Mayer Steven Glazerman

The Effects of Teach for America on Students: Findings from a National Evaluation

Terry Hibpshman Kentucky Education Professional Standards Board June 14, 2004 Paul T. Decker, Daniel P. Mayer, and Steven Glazerman conducted a study, reported in 2004, of the performance of Teach for America teachers as compared to controls, in high-poverty rural and urban schools in several locations. The study compared the mathematics and reading achievement of students of Teach for America teachers and other teachers in the selected schools. An unusual feature for this type of study was that students were randomly assigned to classrooms, removing much of the doubt that often exists about the effect of student and other characteristics on the results. The authors found that students of Teach for America teachers were slightly more proficient in mathematics than those of controls, but that there was no difference in reading achievement.

This is a well-designed study, and the authors in their conclusion seem to imply that it demonstrates the effectiveness of Teach for America teachers relative to teachers recruited from other routes. While the study does show some slight achievement gains due to TFA teachers, the study, taken together with what we know about teaching in high-poverty settings, may not really be indicative of superior performance by TFA teachers. In addition, because of the way that the authors designed the study, and because of features of the TFA recruitment process, the authors may have inadvertently conducted a test of what has become a common belief about effective teaching. To understand how this may have been true, we will begin with a brief discussion of a popular idea about how science proves its theories.

Karl Popper, the twentieth-century philosopher of science, noted that we cannot really ever prove that any of our theories is true. The fact that we have some large number of observations that seem to confirm what we believe says nothing about observations that might be collected in the future, and we might at any moment discover that our cherished beliefs about reality are nothing more than the result of a statistical fluke. Yet we must have some means of deciding whether or not we believe any particular explanation, and he suggested a principle -- namely, that we should subject our theories to "grave risk of refutation" -- i.e., set up experiments that would be unlikely to have the expected outcome if our theories were false. If we do that and the obtained results approximate our expectations, then we have more reason to view our theories as confirmed than we otherwise might.

In the field of teacher quality research and policy, we have heard a number of pronouncements over the past few decades that can be viewed as theoretical statements about what makes for effective teaching. In particular, during the past decade we increasingly have heard that effective teaching results primarily from adequate content knowledge and teacher verbal ability, and that teacher preparation programs that stress "pedagogical" knowledge contribute little or nothing to teacher quality. While there is much discussion of these issues in the research and policy literature, the truth is that there are only a handful of studies that seem to confirm these beliefs, and the results, when one actually reads the studies, are equivocal and relate only to a few academic fields.

Our problem perhaps is that we tend to treat the available research like a set of box scores, which, if we add up the results study by study, could eventually confirm or disconfirm our beliefs. But if we believe Karl Popper's observation on the difficulty of ever proving anything inductively, it is difficult to say how many studies would have to

be done before we were fully convinced – certainly it would be more than have been done thus far.

If we turn the problem on its head, though, we might have at hand a better way of determining the truth or falsity of our theories of teacher effectiveness. Rather than ask whether a given study tends to confirm our beliefs, or whether we have sufficient such studies to make a reasoned judgment about them, we could devise tests of these beliefs that subject them to grave risk of refutation. If the results of such studies fail to confirm our beliefs, then we should have significant doubt that they are true.

The study *The Effects of Teach for America on Students: Findings from a National Evaluation* may be seen in this light. The authors do not propose any particular theory about what constitutes effective teaching or what factors are associated with it, but their findings, in light of what we know about the demographics of students and teachers in high-poverty schools, can be seen as a test, in a Popperian sense, of currently popular beliefs about teachers and teaching.

The beliefs at issue are mentioned above and stated explicitly here as an hypothesis:

H₁: The principal determinants of effective teaching are content knowledge and verbal ability.

If this hypothesis is true, we would expect teachers with greater levels of verbal ability and content knowledge to show markedly better results than teachers with less content knowledge and verbal ability. We would devise an experiment, in a Popperian context, that would compare teachers with high levels of verbal ability and content knowledge with teachers who have lower levels of these traits, and would expect to find striking differences in the outcomes of their students.

Teach for America (TFA), by its nature, makes the assumption that H₁ is true. As the authors of the study ably indicate, these are exactly the criteria for selection of a teacher for the Teach for America program. Candidates are selected from among persons with good academic records, recently graduated from or attending upper-echelon institutions of higher education, who have demonstrated competence in their content areas. Applicants are put through a rigorous screening process designed to select the most motivated and capable candidates, and are then provided with an intensive five-week training process. Those selected for the program are then assigned to schools that have difficulty attracting qualified staff.

It would be difficult to devise a better scenario for subjecting current wisdom on the subject of teacher efficacy to grave risk of refutation. If we test the effectiveness of TFA teachers against the effectiveness of teachers who have less verbal ability and content knowledge, we would expect, given our beliefs that these two factors are essential to effective teaching, that TFA teachers would be much more effective than those with less ability and content knowledge. If we fail to find such results, then we have reason to doubt that the matter is quite so simple. That the study in fact makes this test is attested to by the authors' description of the differences between the TFA and control teachers in the study.

In fact, we find in the results of the study no undeniable confirmation of the verbal ability/content knowledge hypothesis. While the authors do show that TFA

teachers are at least as effective as the control teachers, the truth is that the results are notable for the miniscule differences found between TFA teachers and controls. This is not much support, in a Popperian sense, for H₁. Moreover, as will be seen on closer analysis, the results might well indicate that TFA teachers are not in general as effective as teachers drawn from traditional programs.

To explain why these conclusions might be warranted, we begin by noting a particular strength of this study. Unlike other studies of teacher effectiveness that use statistical controls for such things as school and community effects and student demographics, and always leave at least a bit of doubt about how much the results have been affected by well-known relationships between teachers and the demographics of the students they teach and the schools in which they teach, the authors went to the unusual length of assigning students randomly to classrooms. The between-classroom results are thus directly comparable, without much risk of confounding by student ability or other demographic factors. It is also important to note that the schools selected for the study were fairly homogenous in terms of student demographics, so it is unlikely that the results would have been notably influenced by school or community factors. This is about as pure a design as we are likely to see in the teacher effectiveness literature. Thus, our failure to find results that confirm our expectations tends to leave us with greater doubts about our beliefs than we would have if there were possible confounding relationships in the data.

A final factor in interpreting the results has to do with what we know about teachers and teaching in high-poverty settings. There is ample literature to support the idea that teachers are not randomly assigned to schools. In fact, the more effective and experienced a teacher, the more likely that he/she will migrate to a school in which working conditions are more attractive, and that enrolls students with better levels of ability and lower rates of poverty. This is exactly why the TFA program was developed. Because teachers *can* migrate to more attractive schools, those schools with less capable students and less attractive working conditions have higher turnover rates, have trouble attracting qualified staff, and, as a result, tend to be populated with less-experienced, less-effective teachers. In some states, in fact, it has been found in the past few years that such schools often have very high rates of teachers with emergency credentials or no credentials at all. TFA was developed to attract effective teachers to these problem schools, based on the theory that a pool of such individuals could be found among persons who had not traditionally considered teaching as a career.

The authors used statistical methods that have become fairly standard in teacher-quality research, a value-added approach that regresses teacher characteristics of interest (in this case, whether the teacher had been recruited by TFA) against academic achievement, controlling for previous achievement. Because of the unique design elements of the study, some elements did not have to be considered in the design, and it is likely that the results represent a fairly pure comparison between TFA and control teachers.

The results did find one difference between TFA and control teachers: math score gains showed a very modest improvement for TFA teachers over what could be expected for the controls -- i.e., about one-tenth of a grade level. No difference was found for reading scores.

This is hardly the result we would expect if our beliefs about the predominance of content knowledge and verbal ability as predictors of teacher effectiveness were true. We would expect, given the fact that the control teachers were probably among the least capable in the districts in which the schools were sited, and given the fact that the TFA teachers were, as a group, clearly persons of greater verbal ability than the controls, that there would be a rather large difference between the two groups of teachers. That there was not gives us reason to doubt whether our predictors hold true.

We note here that the principal comparison between the TFA teachers and controls lumped "regularly certified," "less-than-adequately certified," "novice," and "experienced" teachers into a single control group. The authors examined the results separately for "novice" and "experienced" teachers, and found that there was a slightly greater effect size when the TFA teachers were compared with novice teachers, but there was a small positive effect for TFA teachers when compared to either group. If verbal ability and content knowledge were all that mattered, we would expect there to be little or no difference between the effect sizes related to the two comparison groups. Because the effect size declined somewhat from the novice to the more experienced control group, we have some reason to believe that teacher experience and traditional training count for something, even in an unattractive setting in which teachers are probably not the best in the district.

One has to ask, given that the size of the effect declined when TFA teachers were compared to more experienced teachers, what would have happened if a group of TFA teachers had been evaluated in a suburban school setting, with a lower poverty rate, better-motivated students, and teachers self-selected for their demonstrated effectiveness? If the TFA candidates failed to perform much better than poorly trained or untrained teachers in a high-poverty setting, it is conceivable that the effect size would altogether disappear or be reversed if they were compared to better teachers in a more affluent setting.

What can be said about TFA teachers, based on the results of the study, is that they do no worse than the population of teachers who otherwise would be employed by the schools selected for the study. Given that many of the control teachers were less than fully qualified, and given what we know about the demographics of teaching in high-poverty schools, this is faint praise indeed.

While in theory a single failure to confirm an hypothesis may have the effect of eliminating it as a statement about reality, in practice it usually takes several such studies to do so. Observations collected in the course of conducting a study, as well as statistical analyses based on the observations, are sometimes flawed, and further studies may show that the theory is sound but the study itself was inconclusive. We are indebted to the authors of this study for their excellent design, which made it possible to test a popular belief about the relationship between teacher characteristics and effective teaching, but we believe that the final answer on this subject will come only after a number of similar studies have been done.

Policy Implications for Kentucky

The study presents no immediate policy implications for Kentucky. TFA does not serve Kentucky. There is one similar program that serves Jefferson County, but it works

through existing certification programs. In any case, the decision to use TFA and similar programs would probably not be much affected by this study. The study does demonstrate that this type of teacher recruitment program probably presents no more of a problem than the recruitment of emergency-certified staff by public districts

The more significant implications relate to the study's failure to show any strong effect for its teachers. Since we are embarked on an ongoing research program to investigate the relationships between teacher characteristics and student outcomes, it can well serve us as background material to guide us in the development and interpretation of those studies.